

REMARKS

In the patent application, claims 1, 3-41 and 49-56 are pending. In the final office action, all pending claims are rejected.

At section 2 of the final office action, claims 1, 3-42 and 49-56 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. The Examiner states that claims 1, 19, 22, 27, 31 and 32 have the limitation of segmenting audio signals based upon audio characteristics, but it is not clear as to which segmenting aspect of the disclosure this refers. The Examiner states that the specification discloses two aspects of segmenting:

1) a typical audio encoder that extracts audio signal information (outputting segments based upon voice/unvoiced, silence decision denoted as line 110 into the sub-block 12 in Figure 4, generating segmented audio with associated parameters 112 (p.13, lines 8-14 of the specification); and

2) the sub-block 20 re-segments sequence of initial segments based on degree of voicing, etc., derived from speech parameters (Figure 4, p.15, lines 1-17 of the specification).

A. TYPICAL AUDIO ENCODER

It is improper for the Examiner to assert that a typical audio **encoder** outputs segments based upon voice/unvoiced, silence decision without citing a reference or referring to the specification.

As disclosed in the specification, a typical parametric speech coder is used to extract parameters at regular intervals, wherein the parameters include linear prediction coefficients, speech energy (gain), pitch and voicing information (p.11, lines 26-27).

The Examiner errs in stating that the output 112 from the encoder 12 (Figure 4) includes segmented audio with associated parameters.

The Examiner fails to clearly point out where it is disclosed in the specification that the output 112 is segmented audio with associated parameters.

B. SPEECH CODING SYSTEM

It is respectfully submitted that the speech coding system, according to the present invention, includes an encoder block 12 and a compression module 20.

The Examiner errs in pointing to p.13, lines 8-14 of the specification in order to show that the sub-block 12 in Figure 4 generates segmented audio with associated parameters 112 in response to input speech signal 110.

It is respectfully submitted that, p.13, lines 8-14 discloses:

Figure 4 is a speech coding system that quantizes speech parameters 112 utilizing the segmentation information. The compression module 20 can use either quantized parameters from an existing speech coder, or the compression module 20 can use the unquantized parameters directly coming from the parameter extraction unit 12. Moreover, a pre-processing stage (not shown) may be added to the encoder to generate speech signals with specific energy level and/or frequency characteristics. The input speech signal 110 can be generated by a human speaker or by a high-quality TTS algorithm. The encoding of the input speech can be done off-line in a computer, for example.

In the above passage, it is stated that the speech coding system as shown in Figure 4 is used to quantize speech parameters 112 using the segmentation information. Nothing in this passage indicates that the parameters extracted from the input speech signal 110 by the parameter extracting unit 12 is used to partition the audio signal into a plurality of segments.

As disclosed, the speech coding system according to the present invention includes the following units:

- 1) the parameter extraction unit 12 (it is labeled as “encoder” in Figure 4) – this unit 12 receives input signal 110 and outputs parameters 112;
- 2) the compression module 20. Within the compression module 20, there are two sub-units:
 - a software module 22; and
 - a quantizer 24.

As shown in Section D below, segmentation of the input speech signal is carried out in the compression module 20, and quantization of the speech parameters is also carried out in the compression module 20.

C. ENCODER 12

As described in the specification, a typical parametric speech coder is used to extract parameters at regular intervals, including linear prediction coefficients, speech energy, pitch and voicing information. The voicing information can be given as an integer value ranging from 0 to 7, and the parameters can be extracted at 10ms (p.11, lines 26-31). In particular, parameters 112 are extracted from the parameter extraction unit 12, and the parameters 112 can be unquantized (p.13, lines 9-11).

There is no description to show that the sub-block 12 outputs segmented audio with associated parameters. There is no description indicating that the sub-block 12 is used to partition the audio signal 110 into segments.

D. COMPRESSION MODULE 20

The compression module 20 can be used to carry out the following steps:

1. Segmentation of the input speech signal;
2. Definition of the optimal parameter update rate for different segments and parameters;
3. Decimation of transmitted parameters from the original parameters; and
4. Efficient quantization of the derived parameters (p.13, lines 24-28).

It is known in the art that quantization is the process of approximating a continuous range of values (or a very large set of possible discrete values) by a relatively small set of discrete symbols or integer values. Segmentation is the process of dividing of a speech signal into segments at time intervals. According to the present invention, segmentation is based on voicing information as shown in Figures 3a-3d (p.12, line 29 to p.13, line 7). Quantization can be applied to the segments (p.14, lines 24-26). Quantization is also applied to the parameters

(see Figure 5, the quantization block). Thus, “quantizing” and “segmenting” cannot be used interchangeably, as asserted by the Examiner (p.3, fourth paragraph of the office action).

In order to improve the processing efficiency, segmentation of speech signal can be carried out in conjunction with an adaptive downsampling and quantization scheme so that the bit rates and parameter update rates can be adaptively optimized. Such process is carried out in two phases: 1) consecutive frames are divided into continuous segments, and 2) each segment is quantized (p.14, lines 15-26). The first phase process is described on p.14, lines 27-33. The second phase process is carried out by the software module 22 and the quantizer 24 (p.15, lines 1-3). In the flowchart 500 as shown in Figure 7, segmentation takes place at step 513 (p.15, lines 10-13) and the adaptive downsampling and quantization scheme takes place from step 514 to 524 (p.15, lines 13-24). Steps 526 to 530 are only used to repeat the steps 522-524.

It is clear that segmentation and quantization are different processes.

E. 112 REJECTION

As shown in Section D above, segmentation takes place in the compression module 20. As shown in Section C above, the encoder 12 is only used for extracting parameters 112. The Examiner fails to show where the encoder 12 is described as being used for outputting segments, and where, in the specification, the item 112 is described as audio segments.

As shown in Figure 7 and Section D above, segmentation and quantization take place at different steps. The Examiner fails to show where, in the specification, the terms “quantizing” and “segmenting” are used interchangeably.

For the above reasons, the Examiner clearly errs in rejecting claims 1, 3-42 and 49-56 under 35 U.S.C. 112, second paragraph.

F. 102 Rejection Over *Gersho*

At section 3, claims 1, 3-14, 19-21, 26-37, 39-44, 49-56 are rejected under 35 U.S.C. 102(b) as being anticipated by *Gersho et al.* (U.S. Patent No. 6,311,154, hereafter referred to as *Gersho*).

In rejecting claim 1, the Examiner states that *Gersho* teaches segmenting {partitioning or classifying} the audio signal {speech} into a plurality of segments {frames} (partitioning

samples of a speech signal into frames, col. 4, lines 25-27) based on the audio characteristics {classes} of the audio signal (classifying the speech signal in each from into one of a plurality of classes, col. 4, lines 25-27); and encoding the segments {frames} with different encoding settings {excitation} (encoding an excitation for the frame using one of a plurality of excitation coding...selected according to the class of the frame, col. 4, lines 30-33).

F.1 The Claimed Invention

It is respectfully submitted that claim 1 includes the limitations of
 obtaining, for each of a plurality of consecutive time intervals, one or more parameters from an audio signal, said one or more parameters indicative of audio characteristics of the audio signal,
 partitioning the audio signal into a plurality of segments based on the parameters obtained for the consecutive time intervals; and
 encoding the segments with different encoding settings.

Because “partitioning” is based on the parameters, “partitioning” must be carried out after the parameters are obtained.

In *Gersho*, the partitioning of the speech signal into frames is carried out before classification of the speech signal in the frame, as shown in sub-section F.2(a) below.

F.2 The Cited *Gersho* Reference

The Examiner states that *Gersho* teaches segmenting {partitioning or classifying} the audio signal {speech} into a plurality of segments {frames} (partitioning samples of a speech signal into frames, col. 4, lines 25-27) based on the audio characteristics {classes} of the audio signal (classifying the speech signal in each from into one of a plurality of classes, col. 4, lines 25-27).

In col.4, lines 23-34, *Gersho* discloses:

Further in accordance with this invention there is provided a method for coding a speech signal that includes steps of (a) partitioning samples of a speech signal into frames; (b) deriving a residual signal for each frame; (c) classifying the speech signal in each frame into one of a

plurality of classes; (d) identifying the location of at least one window in the frame by examining the residual signal for the frame; (e) encoding an excitation for the frame using one of a plurality of excitation coding techniques selected according to the class of the frame; and, for at least one of the classes, (f) confining all or substantially all of non-zero excitation amplitudes to lie within the windows.

In the above passage, *Gersho* discloses a method in which the speech signal containing speech samples is partitioned or segmented into frames in step (a) and the speech signal in each frame is classified into one of the classes in step (c). As shown in Figure 8, *Gersho* uses a high-pass filtering block 30 to filter the input speech into high-pass filtered speech (col.14, lines 56-57). The high-pass filtered speech is divided into non-overlapping “frames” of 160 samples each (col.15, lines 7-8). According to *Gersho*, all the frames contain the same number of samples. The filtered speech is provided to two separate modules: parameter estimation module 32 and open-loop classifier module 34. The residual signal and the model parameters for each of the frames are extracted or estimated by the model parameter estimation module 32. The classification information OLC(m) is provided to an excitation encoding and speech synthesis module 42. The residual signal and parameters, along with the classification information, are used for speech synthesis.

F.2(a) Partitioning and Classification according to *Gersho*

The Examiner errs in stating that *Gersho* teaches segmenting the audio signal into a plurality of segments based on the audio characteristics of the audio signal.

According to *Gersho*, the partitioning of the speech signal in step (a) is carried out in the High-pass filtering block 30 and classification of the speech samples in step (c) is carried out in the Open-loop classifier 34. Also, a Model Parameter Estimation block 32 is used to obtain various parameters (col.15, lines 8-24). Since the Open-loop classifier 34 and the Model Parameter Estimation block 32 receives the filtered speech from the High-pass filtering block 30, the partitioning of the speech signal in step (a) cannot be based on the classification of speech samples on the parameters.

In *Gersho*, partitioning of the speech signal into frames is carried out before the classification of the speech signal in the frame.

Gersho does not disclose partitioning the audio signal into a plurality of segments based on the parameters obtained for the consecutive time intervals.

F.2(b) Segmentation and Classification

The Examiner errs in equating segmenting with classifying.

According to *Gersho*, segmentation is carried out using a high-pass filtering block 30, whereas classification is carried out in a separate open-loop classifier 34 (Figure 8). In particular, the high-pass filtering block 30 is used to filter the input speech into high-pass filtered speech (col.14, lines 56-57). The high-pass filtered speech is divided into non-overlapping “frames” of 160 samples each (col.15, lines 7-8). The open-loop classifier unit 34 determines the nature (voiced, unvoiced or transition) of the speech in frame and the output of the classifier in frame can be UNVOICED or NOT UNVOICED. This frame-level decision is combined with the subframe-level decisions and then outputted from the open-loop classifier unit 34 (col.16, lines 5-23).

There is no description in *Gersho* indicating that segmenting is the same as classifying.

F.3 102 Rejection of Claim 1

As pointed out in sub-section F.2(a) above, *Gersho* does not disclose or suggest partitioning the audio signal into a plurality of segments based on the parameters obtained for the consecutive time intervals.

For the above reasons, *Gersho* fails to anticipate claim 1.

F.4 102 Rejection of Claims 19 and 27

In rejecting claims 19 and 27, the Examiner states that *Gersho* teaches an input for receiving audio data indicative of parameters in the adjusted representation (input applied to element 14, Figure 3) and a module responsive to the audio data for generating the audio signal based on the adjusted signals and the characteristics of the audio signal (Figure 3).

The Examiner states that it would have been inherent to one of ordinary skill in the art to use a decoder in order to reverse the encoding data for further processing, such as modulating or storing the audio signal.

F.4(a) Claimed Invention

It is respectfully submitted that claim 19 is directed to a decoder which comprises:

an input for receiving audio data indicative of a plurality of segments of an audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal, and wherein the plurality of segments are obtained by partitioning the audio signal based on the parameters extracted for the consecutive time intervals, and the audio data is indicative of the parameters in an adjusted representation; and

a module, responsive to the audio data, for generating a further audio signal based on the adjusted representation and the encoding settings.

Claim 27 is directed to an electronic device comprising:

an input module for receiving audio data indicative of a plurality of segments of an audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters indicative of audio characteristics of the audio signal, and wherein the plurality of segments are obtained by partitioning the audio signal based on the parameters extracted for the consecutive time intervals, and the audio data is indicative of the parameters in an adjusted representation; and

a decoder, responsive to the audio data, for generating a synthesized audio signal based on the adjusted representation.

F.4(b) *Gersho* does not disclose a decoder as claimed

In Figure 3, *Gersho* discloses a circuit for obtaining smoothed energy contour of a speech residual signal (col. 5, lines 38-40) in a speech encoder 12 (col.7, lines 18-23). In particular, the speech encoder 12 has a fixed frame structure (basic frame structure) and each basic frame is partitioned into M equal length subframes (col.7, lines 18-27). In conventional AbS coding schemes, the excitation signal for each subframe is selected by a search operation. However, an adequate precise representation of the excitation segment is difficult to obtain (col.7, lines 28-33). *Gersho* discloses a method for identifying the location of the excitation activity into the subframe by examining a smoothed energy contour of the linear prediction residual (col.7, lines 34-50). Figure 3 depicts an encoder wherein a linear prediction (LP) whitening filter 14 for

forming the residual signal from the input speech in order to obtain a smooth energy contour and to identify the energy peaks (col. 8, line 53 to col.9, line 2).

In Figure 3, the block 14 is a filter for obtaining a residual signal by filtering an input speech in a speech encoder.

Thus, Figure 3 has nothing to do with a decoder.

F.4(c) Inherent Use of Decoder

The Examiner states that, at the time of the invention, it would have been inherent to one of ordinary skill in the art to use a decoder in order to reverse the encoding data for further processing, such as modulating or storing the audio signal.

The Examiner fails to clearly explain how the filter 14 in Figure 3 is related to a decoder that is used to reverse the encoding data.

The Examiner fails to point out where *Gersho* discloses a decoder comprising an input for receiving audio data indicative of a plurality of segments of an audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal, and wherein the plurality of segments are obtained by partitioning the audio signal based on the parameters extracted for the consecutive time intervals, and the audio data is indicative of the parameters in an adjusted representation.

F.5 *Gersho* Fails to Anticipate Claims 19 and 27

As pointed out in sub-section F.4(c) above, *Gersho* does not disclose or suggest partitioning the audio signal into a plurality of segments based on the parameters extracted for the consecutive time intervals.

As pointed out in sub-section F.4(b) above, *Gersho* does not disclose a decoder as claimed.

For the above reason, *Gersho* fails to anticipate claim 19 and 27.

F.6 *Gersho* Fails to Anticipate Claim 31

In rejecting claim 31, the Examiner states that *Gersho* discloses implementing a cell phone system (col. 6, lines 33-36).

It is respectfully submitted that claim 31 includes the limitation of an input module for receiving audio data from at least one of the base stations, the audio data indicative of a plurality of segments of an input audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal, and wherein the plurality of segments are obtained by partitioning the input audio signal based on the parameters extracted for the consecutive time intervals and encoded with a plurality of encoding settings based on the audio characteristics, the audio data indicative of the parameters in an adjusted representation.

As with claims 19 and 27 above, *Gersho* does not disclose an input for receiving audio data indicative of a plurality of segments of an audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal, and wherein the plurality of segments are obtained by partitioning the audio signal based on the parameters extracted for the consecutive time intervals, and the audio data is indicative of the parameters in an adjusted representation.

For the above reasons, *Gersho* fails to anticipate claim 31.

G. 102 Rejection Over *Sinha*

At section 5, claims 15-18, 22-25, 38, 45 are rejected under 35 U.S.C. 102(e) as being anticipated by *Sinha et al.* (U.S. Patent No. 7,191,136, hereafter referred to as *Sinha*)

G.1 The Claimed Invention

It is respectfully submitted that claim 22 includes the limitations of
 an input for receiving audio data indicative of parameters obtained from an audio signal in a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal; and

an adjustment module for adjusting one or more of the parameters for providing an adjusted representation of the parameters, wherein said adjusting comprises partitioning the audio signal into a plurality of segments based on the parameters obtained for the consecutive

time intervals and encoding the segments based on one or more of a plurality of encoding settings.

Thus, claim 22 includes the limitation of partitioning the audio signal into a plurality of segments based on the parameters obtained for the consecutive time intervals.

G.2 102 Rejection of Claim 22

In rejecting claim 22, the Examiner states that *Sinha* discloses a method and device for encoding as claimed (col.4, lines 47-51; col.3, lines 1-6; col.6, lines 43-47 and col.7, lines 42-44). In particular, the Examiner states that

Sinha teaches a method for use in a parameter audio coding to encode an audio signal by segmenting the audio signal, for each of a plurality of consecutive time intervals, one or more parameters from an audio signal, said one or more parameters relating to audio characteristics of the audio signal (col.4, lines 47-51, by high pass filtering the input audio signal); performing a non-linear parameter representation of the signal (col. 4, lines 53-59 – wherein the data amount per processing depends upon the frequency characteristics of the audio signal, and the characteristics analyzed can be peak analysis, lattice quantization, or frequency range selection – col.3, lines 1-6); and encoding the segments with different encoding settings (by choosing compression settings on-the-fly – col.6, lines 43-47)

It seems that the Examiner considers the non-linear parameter representation of the signal as being equivalent to the parameters relating to audio characteristics of the audio signal as claimed. However, the Examiner fails to clearly point out what the Examiner considers as being equivalent to the plurality of segments partitioned from the audio signal based on the parameters obtained for the consecutive time intervals.

It is respectfully submitted that while parameters are obtained from the audio signal, they are not segments of the audio signals.

G.3 The *Sinha* Reference

Sinha discloses a method for improving an audio compression scheme, such as perceptual audio coding (PAC). In a conventional PAC scheme, as shown in Figure 1, the input signal is

segmented into frames to be stored in a frame buffer 104. The frames are then processed through a long-term predictor 106 and a short-term predictor 108 for linear predictive analysis (col. 2, lines 13-16). Each of the audio frames in PAC consists of 1024 pulse code modulated (PCM) samples (col.3, lines 52-56). According to *Sinha*, as the input speech signal is segmented into frames of 1024 PCM samples, the speech signal is simultaneously provided to a low-pass filter 402 for obtaining compressed information consisting of coded low frequency components, and to a high-pass filter 404 for obtaining a parametric representation of the high frequency components based on a non-linear model 406. The parametric representation is updated every audio frame in order to estimate the non-linear parameters 408 (col.4, lines 43-59).

According to *Sinha*, segmentation of the speech signal into frames takes place before the high-pass and low-pass filtering and before the non-linear parameters 408 are obtained. Furthermore, the parametric representation is updated every audio frame.

Thus, *Sinha* does not disclose or suggest partitioning the audio signal into a plurality of segments based on the parameters obtained for the consecutive time intervals.

For the above reasons, *Sinha* fails to anticipate claim 22.

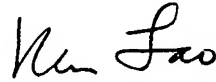
H. Dependent Claims

As for claims 3-18, 20-21, 23-26, 28-30, 32-41 and 49-56 they are dependent from claims 1, 19, 27 and 31 and include further limitations. For reasons regarding claims 1, 19, 27 and 31 above, *Gersho* or *Sinha* also fails to anticipate claims 3-18, 20-21, 23-26, 28-30, 32-41 and 49-56.

CONCLUSION

Claims 1, 3-41 and 49-56 are allowable. Early allowance of claims 1, 3-41 and 49-56 is earnestly solicited.

Respectfully submitted,



Kenneth Q. Lao

Attorney for the Applicant
Registration No. 40,061

Date: *September 29, 2010*

WARE, FRESSOLA, VAN DER SLUYS
& ADOLPHSON LLP
Bradford Green, Building Five
755 Main Street, P.O. Box 224
Monroe, CT 06468
Telephone: (203) 261-1234
Facsimile: (203) 261-5676
USPTO Customer No. 004955